

Simultaneous adjustment of bias and coverage probabilities for confidence intervals

P. MENÉNDEZ¹, Y. FAN¹, P. H. GARTHWAITE² and S. A. SISSON¹

October 15, 2012

ABSTRACT

We give a method for the correction of confidence intervals when the original interval does not have the correct nominal coverage probabilities in the frequentist sense. Our method is general and does not require any distributional assumptions. It can be applied to both frequentist and Bayesian inference where interval estimates are desired. We provide theoretical results for the consistency of our proposed estimator, and provide two complex examples, on confidence interval correction for composite likelihood estimators and in approximate Bayesian computation, to demonstrate the wide applicability of our method.

Keywords: Approximate Bayesian computation; Confidence interval correction; Composite likelihood; Coverage probability.

1 Introduction

Interval estimates are typically intended to have a specified level of coverage. This is true, for example, of both frequentist confidence intervals and Bayesian credible intervals. However, for many problems the coverage of a confidence or credible interval will only equal its nominal value asymptotically, and coverage can be poor even for quite large samples in some situations. In many complex problems, there can be inherent bias which can be difficult to quantify or calculate. This can arise, for example, in composite likelihood problems (Varin et al. 2011) and approximate Bayesian computation (Sisson and Fan 2011). In this paper we propose a novel procedure for adjusting interval estimates that has wide application and will typically reduce the bias in their coverage.

The procedure assumes that the mechanism that generated the sample data could be simulated if population parameters were known. These parameters are estimated by sample statistics derived from real data, and then pseudo-samples are drawn from the estimated population distribution. From each pseudo-sample a confidence/credible interval is determined for the quantity of interest. The frequentist bias in these intervals is calculated and then used to adjust the interval estimate given by the real data.

¹School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, AUSTRALIA

²Department of Mathematics and Statistics, Open University, Milton Keynes, MK7 6AA, U.K.

The method has similarities to the double-bootstrap (Davison and Hinkley 1997), in which a bias correction is applied to a bootstrap interval by re-sampling from a bootstrap distribution. A difference in our method is that it involves only one level of sample generation, which makes it computationally less demanding, although the saving may dissipate if computationally demanding methods (such as Markov chain Monte Carlo) are used to obtain the interval estimates from sample data.

In some of our examples, we apply our procedure to reduce bias in the coverage of Bayesian credible intervals. This may not seem intuitive, since *coverage* is a frequentist property while a Bayesian interval may reflect personal probabilities. However, there are many situations where posterior distributions should preferably be well calibrated. These include inference with objective or probability matching prior distributions, the verification of Bayesian simulation software (Cook et al. 2006) and techniques and diagnostics in likelihood-free Bayesian inference (Fearnhead and Prangle 2012; Prangle et al. 2012).

In Section 2 we describe the proposed method and give theoretical results related to it, illustrating them through simulated examples. In Section 3 we apply our method to two more complex, real analyses. One of these involves estimation with composite likelihoods, which is known to produce confidence intervals that are too narrow, and the other involves approximate Bayesian computation, which typically gives larger posterior credibility intervals than desired. Some concluding comments are given in Section 4.

2 Coverage correction for confidence intervals

Suppose we are interested in estimating an equal-tailed $100(1 - \alpha)\%$ confidence interval for some parameter $\theta \in \Theta \subseteq \mathcal{R}$. Thus for observed data \mathbf{x} , we seek an estimate $L(\mathbf{x})$, such that

$$P(\theta \leq L(\mathbf{x})) = \alpha/2,$$

where $L(\mathbf{x})$ denotes the lower limit of the interval. Similarly for the upper limit, we seek an estimate $U(\mathbf{x})$, such that,

$$P(\theta \geq U(\mathbf{x})) = \alpha/2.$$

In the frequentist setting, the expressions above are written in terms of pivotal functions of the data, \mathbf{x} , and parameter, θ , since the parameter θ is considered a fixed quantity. In the Bayesian setting, credible intervals are based on the posterior distribution of θ . In an abuse of notation, we will use the above notation in both cases. For a given θ , coverage probabilities can be estimated by simulating from the data model $f(\mathbf{x}|\theta)$ multiple times, constructing the relevant intervals, and then counting the proportion of intervals which contain θ .

We will assume that the method of obtaining the lower and upper intervals $L(\mathbf{x})$ and $U(\mathbf{x})$ is known, but this can be generic. We do not assume that these estimates produce the correct coverage probability. We also assume that the population parameters can be well approximated from the data. Our goal is to provide a method that gives adjustments that improve this coverage. We first give theoretical results for the proposed methodology, and then give details of its implementation.

2.1 Theoretical results

Assumption 1 We suppose that the observed data \mathbf{x} come from the model given by $f(\mathbf{x}|\theta)$, $\theta \in \Theta$. For any $\theta \in \Theta$, we assume that it is possible to simulate from $f(\cdot|\theta)$.

Assumption 2 Given θ and data $\mathbf{x} \sim f(\mathbf{x}|\theta)$ there exists a consistent estimator $\tilde{\theta}$ of θ .

Assumption 1 requires that we are able to simulate replicate data from the model given the values of the parameters. Assumption 2 requires that we have a good estimator for θ , so that interval estimates obtained using $\tilde{\theta}$ converge to those estimates obtained using the population parameter θ , as the amount of data gets large.

In the following, we only require the lengths of the intervals to be consistent. Consequently, Assumption 2 is not always necessary. For example, if θ represents a location parameter whose confidence interval has a length that is independent of θ (see later example). In the frequentist setting, the maximum likelihood estimator of θ is consistent and unbiased in many finite sample situations. In the Bayesian setting, the posterior distribution is consistent under mild assumptions, and the posterior mean estimate of θ is asymptotically unbiased. However, in both cases, finite sample bias in $\tilde{\theta}$ may render our method less accurate.

Theorem 2.1 For some θ and $\mathbf{x} \sim f(\mathbf{x}|\theta)$, let $L(\mathbf{x})$ be an estimator of the lower limit of a $100(1 - \alpha)\%$ level confidence interval, and suppose that

$$P\{\theta < L(\mathbf{x})\} \neq \alpha/2.$$

Let $G_{\{W\}}$ denote the distribution function of a random variable W . Consider the new estimator

$$L_c(\mathbf{x}) = L(\mathbf{x}) + \xi_{\alpha/2}, \tag{1}$$

where $\xi_{\alpha/2}$ is the $\alpha/2$ -th quantile of the distribution function $G_{\{\theta - L(\mathbf{x})\}}$, so that $G_{\{\theta - L(\mathbf{x})\}}(\xi_{\alpha/2}) = \alpha/2$. Then the new estimator, $L_c(\mathbf{x})$, will have the correct coverage probability

$$P\{\theta < L_c(\mathbf{x})\} = \alpha/2.$$

Proof: See Appendix.

From the above theorem, it can then be seen that for the estimator of the upper limit of a $100(1 - \alpha)\%$ confidence interval, $U(\mathbf{x})$, we can write

$$U_c(\mathbf{x}) = U(\mathbf{x}) + \xi_{1-\alpha/2}, \tag{2}$$

where $\xi_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the distribution function $G_{\{\theta - U(\mathbf{x})\}}$, so that $G_{\{\theta - U(\mathbf{x})\}}(\xi_{1-\alpha/2}) = 1 - \alpha/2$. In this case, we then have that

$$P\{\theta > U_c(\mathbf{x})\} = \alpha/2.$$

Theorem 2.2 For some θ and observed data $\mathbf{x} \sim f(\mathbf{x}|\theta)$, suppose the lower limit of a $100(1-\alpha)\%$ confidence interval $L(\mathbf{x})$ can be obtained, and that this estimate does not necessarily give the correct coverage probability. Suppose that $\tilde{\theta} \in \mathcal{R}$ is a consistent estimator of θ , evaluated using the data \mathbf{x} . Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be n replicate datasets simulated independently from $f(\cdot|\tilde{\theta})$, and denote the corresponding lower confidence limits by $L_1(\mathbf{y}_1), \dots, L_n(\mathbf{y}_n)$, obtained in the same manner as $L(\mathbf{x})$. Define

$$\hat{G}_{\{\tilde{\theta}-L(\mathbf{y})\}}(\epsilon) = \frac{1}{n} \sum_{i=1}^n I_{\{\tilde{\theta}-L_i(\mathbf{y}_i) < \epsilon\}}$$

as the empirical distribution of $\tilde{\theta} - L(\mathbf{y})$ based on the observed values of $\tilde{\theta} - L_i(\mathbf{y}_i)$, $i = 1, \dots, n$. If we define

$$\tilde{L}_c(\mathbf{x}) = L(\mathbf{x}) + \hat{\xi}_{\alpha/2} \quad (3)$$

where $\hat{\xi}_{\alpha/2} = \hat{G}_{\{\tilde{\theta}-L(\mathbf{y})\}}^{-1}(\alpha/2)$, then $\tilde{L}_c(\mathbf{x})$ is a consistent estimator of $L_c(\mathbf{x})$, as defined in Equation (1).

Proof: See Appendix.

In combination, Theorems 2.1 and 2.2 state that if we simulate data $\mathbf{y}_1, \dots, \mathbf{y}_n \sim f(\mathbf{y}|\tilde{\theta})$ and subsequently obtain the confidence limits $L_1(\mathbf{y}_1), \dots, L_n(\mathbf{y}_n)$ in the same way as for the original data \mathbf{x} , then we can correct the bias in the original lower limit estimate, $L(\mathbf{x})$, by addition of the $\alpha/2$ -th sample quantile of $\tilde{\theta} - L_1(\mathbf{y}_1), \dots, \tilde{\theta} - L_n(\mathbf{y}_n)$.

Corollary 1 Under the assumptions in Theorem 2.2, a central limit theorem holds for $\tilde{L}_c(\mathbf{x})$. Specifically, for all $\alpha \in (0, 1)$, $\theta \in \mathcal{R}$ and $\mathbf{x} \sim f(\mathbf{x}|\theta)$, we have that

$$\sqrt{n}(\tilde{L}_c(\mathbf{x}) - L_c(\mathbf{x}))G'_{\{\theta-L(\mathbf{x})\}}(\xi_\alpha) \longrightarrow N(0, \alpha(1-\alpha))$$

as $n \rightarrow \infty$, where $G'_{\{\theta-L(\mathbf{x})\}}(\xi) = \frac{\partial}{\partial \xi} G_{\{\theta-L(\mathbf{x})\}}(\xi)$, and ξ_α is the α -th quantile of $G_{\{\theta-L(\mathbf{x})\}}$.

Proof: The result follows immediately from Equation (7) of the proof for Theorem 2.2 (see Appendix).

The above theoretical results provide a simple way of estimating corrections to the lower and upper confidence limits that will produce the correct nominal coverage probability. In addition, these estimators are consistent and asymptotically normal.

2.2 Correction procedure

In summary, the correction algorithm has the following steps:

Step 1 Obtain $L(\mathbf{x})$ and $U(\mathbf{x})$, the upper and lower limits of the desired $100(1-\alpha)\%$ confidence interval for the parameter θ , for an observed dataset \mathbf{x} .

Step 2 Evaluate $\tilde{\theta}$ and generate n independent datasets $\mathbf{y}_1, \dots, \mathbf{y}_n \sim f(\mathbf{y}|\tilde{\theta})$ from the model.

Step 3 For each dataset \mathbf{y}_i , compute the $100(1-\alpha)\%$ lower and upper confidence limits, $L_i(\mathbf{y}_i)$ and $U_i(\mathbf{y}_i)$, for the parameter $\tilde{\theta}$, using the same method as in Step 1.

Step 4 Set the corrected lower and upper limits to

$$\begin{aligned}\tilde{L}_c(\mathbf{x}) &= L(\mathbf{x}) + \hat{G}_{\{\tilde{\theta} - L(\mathbf{y})\}}^{-1}(\alpha/2) \\ \tilde{U}_c(\mathbf{x}) &= U(\mathbf{x}) + \hat{G}_{\{\tilde{\theta} - U(\mathbf{y})\}}^{-1}(1 - \alpha/2)\end{aligned}$$

where $\hat{G}_{\{W\}}^{-1}(\alpha)$ denotes the α -th sample quantile of the random variable W .

2.3 Simple examples

We illustrate the above procedure with two simple examples. In the first, we consider confidence interval correction for the mean parameter of a normal distribution with known variance. In the second example, it is assumed that the mean is known and that we are interested in the variance parameter.

Example 1: Normal distribution with known variance

Suppose that θ is the location parameter of a Normal distribution with unit variance, so that $x_i \sim N(\theta, 1)$ where $\mathbf{x} = (x_1, \dots, x_m)$. In this case, the maximum likelihood estimator is $\tilde{\theta} = \bar{x} = \sum_i x_i/m$. For illustration, we suppose that the confidence interval we obtain for θ does not have the correct coverage, in that we obtain the equivalent confidence interval when data are generated from $x_i \sim N(\theta, (1 + \epsilon)^2)$ with $\epsilon \geq 0$. The value of ϵ controls the amount of error in the coverage probability. Following the usual frequentist approach, the $100(1 - \alpha)\%$ confidence interval for θ is given by $L(\mathbf{x}) = \bar{x} - z_{\alpha/2}(1 + \epsilon)/\sqrt{m}$ and $U(\mathbf{x}) = \bar{x} + z_{1-\alpha/2}(1 + \epsilon)/\sqrt{m}$, where z_α is the α -th quantile of the standard normal distribution. Clearly the correction for the interval when $\epsilon > 0$ is $L_c(\mathbf{x}) = L(\mathbf{x}) + z_{\alpha/2}\epsilon/\sqrt{m}$ and $U_c(\mathbf{x}) = U(\mathbf{x}) - z_{1-\alpha/2}\epsilon/\sqrt{m}$.

Figure 1 displays the results of the correction procedure for a 95% confidence interval based on 100 replicate analyses. Each analysis is based on samples of size $m = 20$ with $\theta = 0$, so that $x_1, \dots, x_m \sim N(0, 1)$, and $n = 100$ replicated samples $\mathbf{y}_1, \dots, \mathbf{y}_n$ with elements drawn from $N(\tilde{\theta}, 1)$. Figure 1 (top plots) illustrates the corrected confidence limits $\tilde{L}_c(\mathbf{x})$ and $\tilde{U}_c(\mathbf{x})$ for a range of error term values, ϵ . Clearly the correction produces an unbiased adjustment, as the boxplots are centred on the true confidence bounds (the horizontal line) in each case. Further, the performance of the method produces qualitatively the same corrected interval limits, irrespective of the value of ϵ .

The bottom plots display the corrections $\tilde{L}_c(\mathbf{x})$ and $\tilde{U}_c(\mathbf{x})$ with $\epsilon = 1$ fixed, for a range of values of $\tilde{\theta}$. For this example, choosing $\tilde{\theta}$ to be any arbitrary value will result in the same quality of unbiased correction. This arises as the distributions of $\tilde{\theta} - L(\mathbf{y})$ and $\tilde{\theta} - U(\mathbf{y})$ do not change with $\tilde{\theta}$, so that the confidence intervals all have the same width as $\tilde{\theta}$ varies. As this is a location parameter only analysis, this is one case where Assumption 2 is not required to produce a consistent adjustment (see Section 2.1).

Example 2: Normal distribution with known mean

Suppose now that θ is the scale (variance) parameter of a Normal distribution with mean zero, so that $x_i \sim N(0, \theta)$. Here we specify $\tilde{\theta} = S^2 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2$ as the sample variance. In this setting, suppose that the regular confidence limits for θ are biased

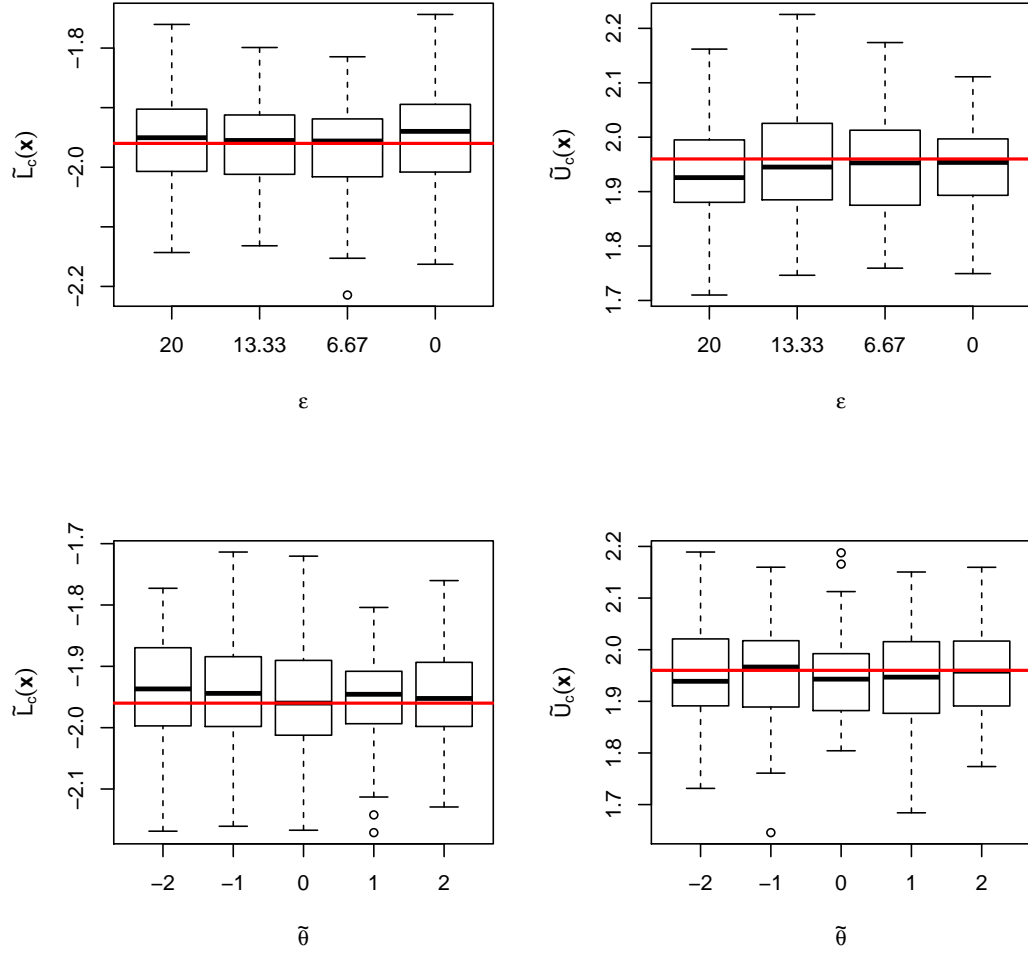


Figure 1: Lower $\tilde{L}_c(\mathbf{x})$ and upper $\tilde{U}_c(\mathbf{x})$ corrected confidence limit estimates for 100 replicated analyses for the normal location model. Top plots show the corrected limits for $\epsilon = 20, 13.33, 6.67, 0$ with $\tilde{\theta} = \bar{x}$. Bottom plots show the corrected limits for $\tilde{\theta} = -2, -1, 0, 1, 2$ with $\epsilon = 1$. The horizontal lines represent the 0.025-th (left plots) and 0.975-th (right plots) percentiles of a standard normal distribution.

downwards by a constant value $\epsilon > 0$. Specifically, the $100(1 - \alpha)\%$ confidence interval for θ is given by $L(\mathbf{x}) = \frac{(m-1)S^2}{\chi_{1-\alpha/2; m-1}^2} - \epsilon$ and $U(\mathbf{x}) = \frac{(m-1)S^2}{\chi_{\alpha/2; m-1}^2} - \epsilon$, where $\chi_{\alpha, k}^2$ denotes the α -th percentile of a χ_k^2 distribution with k degrees of freedom.

Figure 2 shows the results of the correction procedure for the lower limit of a 95% confidence interval based on 100 replicate analyses. Each analysis uses samples of size m with $\theta = 1$, so that $x_1, \dots, x_m \sim N(0, 1)$, and $n = 2000$ replicated samples $\mathbf{y}_1, \dots, \mathbf{y}_n$ with elements drawn from $N(0, \tilde{\theta})$. Figure 2 (left panel) illustrates the corrected lower confidence limit, $\tilde{L}_c(\mathbf{x})$, based on a sample of size $m = 20$, for a range of fixed values of $\tilde{\theta}$. The extreme left and right boxplots correspond to the raw biased ($L(\mathbf{x})$) and true unbiased ($L_c(\mathbf{x})$) limits respectively. Clearly, as $\tilde{\theta}$ changes, then so does the location of the adjusted limits. This occurs as, in contrast with the above example, the distributions of $\tilde{\theta} - L(\mathbf{y})$ and $\tilde{\theta} - U(\mathbf{y})$ clearly do change with $\tilde{\theta}$. When $\tilde{\theta} = \theta = 1$, then the correction procedure produces the correct adjusted limits, as indicated by the rightmost boxplot. Hence, it is necessary to use the right value for $\tilde{\theta}$ when making the correction.

Assumption 2 requires that $\tilde{\theta}$ is a consistent estimator of θ . Hence we can be sure that $\tilde{\theta} \rightarrow \theta$ as $m \rightarrow \infty$, and as a result that the distribution of $\tilde{\theta} - L(\mathbf{y})$ approaches that of $\theta - L(\mathbf{x})$, so that our correction procedure will perform correctly for large enough m . In practice, the required value of m can be moderate. Figure 2 (right panel) shows how the correction error, $\tilde{L}_c(\mathbf{x}) - L_c(\mathbf{x})$, varies as a function of m . Clearly, the median error is close to zero even for small sample sizes. However, there is some asymmetry for small m , which is also visible in the left panel (e.g. compare the differences in the bias in the boxplots with $\tilde{\theta} = 0.4$ and $\tilde{\theta} = 1.6$), although this is eliminated as m increases.

3 Real examples

We now consider interval estimation in two real, complex modelling situations. The first is an application of composite likelihood techniques in the modelling of spatial extremes. With composite likelihoods, deriving unbiased confidence intervals can require a large amount of algebra, whereas biased intervals that are typically too narrow are easily computable. The second is an application of approximate Bayesian computation (ABC) methods in the modelling of a time series of g -and- k distributed observations. In most practical settings the mechanism behind the model fitting process within the ABC framework typically gives posterior credible intervals that are too large.

3.1 Spatial extremes via composite likelihoods

In the context of analysing spatial extremes, Padoan et al. (2010) developed a pairwise composite likelihood model, for inference using max-stable stationary processes. Specifically, for m annual maximum daily rainfall observations, at each of K spatial locations, the pairwise composite likelihood was specified as

$$\ell_C(\theta|\mathbf{x}) = \sum_{i < j} w_{ij} \log f(\mathbf{x}_i, \mathbf{x}_j|\theta)$$

where $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_K)$ and $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$, $f(\mathbf{x}_i, \mathbf{x}_j|\theta)$ is a known bivariate density function with parameter vector θ evaluated at spatial locations i and j , and $w_{ij} > 0$ are

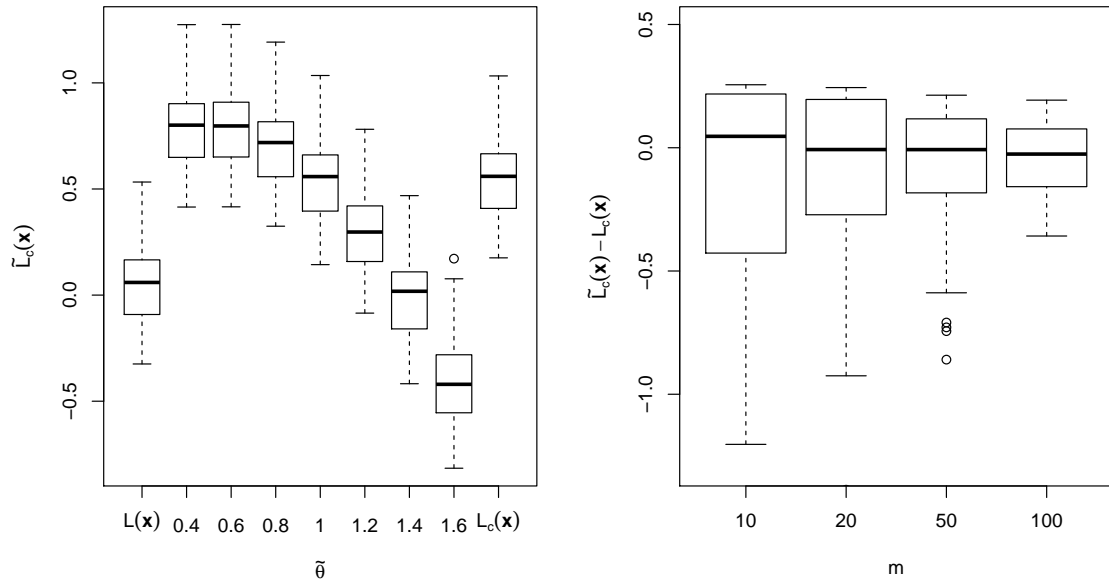


Figure 2: Left panel: Boxplots of the corrected lower confidence limit, $\tilde{L}_c(\mathbf{x})$, when holding $\tilde{\theta}$ fixed at various values $\tilde{\theta} = 0.4, \dots, 1.6$ (true value is $\tilde{\theta} = 1$). Leftmost and rightmost boxplots correspond to the biased ($L(\mathbf{x})$) and true unbiased ($L_c(\mathbf{x})$) lower limits respectively. Right panel: Boxplots of the correction error $\tilde{L}_c(\mathbf{x}) - L(\mathbf{x})$ as a function of observed data sample size m . All boxplots are based on 100 replicate analyses.

weights such that $\sum_{i,j} w_{ij} = 1$. Under the usual regularity conditions, the maximum composite likelihood estimator, $\hat{\theta}$ can provide asymptotically unbiased and normally distributed parameter estimates when standard likelihood estimators are unavailable (e.g. Varin et al. 2011).

Specifically, we have (e.g. Huber 1967) that $\hat{\theta} \sim N(\theta, \tilde{I}^{-1}(\hat{\theta}))$, with

$$I(\theta) = H(\theta)J(\theta)^{-1}H(\theta). \quad (4)$$

where $H(\theta)$ and $J(\theta)$ are respectively the expected information matrix and the covariance matrix of the score vector. In the ordinary maximum likelihood setting, $H(\theta) = J(\theta)$. In the max-stable process framework, Padoan et al. (2010) provided an analytic expression for $J(\theta)$ for a particular (Gaussian) spatial dependence model. Combined with the standard numerical estimates of $H(\theta)$, this allowed for the construction of standard confidence intervals for θ . However, for composite likelihood techniques in general, obtaining analytic expressions or numerical estimates of $J(\theta)$ can be challenging, whereas estimates of $H(\theta)$ are readily available. In this example, we demonstrate how our proposed method can be employed to correct the too narrow confidence intervals that result from using $I(\theta) = H(\theta)$. We then compare our results with those derived from the known maximum composite likelihood information matrix (4).

We considered four spatial models for stationary max-stable processes that describe different degrees of extremal dependence, with parameter inference based on $m = 100$ observations at each of $K = 50$ randomly generated spatial locations. Each model expresses the degree of extremal dependence via the covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$

where the values for each parameter for each model M_1, \dots, M_4 are given in Table 1. Model M_4 has an additional non-stationary spatial component that is modelled by the marginal parameters μ, λ and ξ (corresponding to location, scale and shape parameters) through the response surface

$$\begin{aligned} \mu &= \alpha_0 + \alpha_1 * lat + \alpha_2 * lon \\ \lambda &= \beta_0 + \beta_1 * lon \\ \xi &= \gamma_0, \end{aligned}$$

where lat and lon denote latitude and longitude coordinates.

Model	σ_1^2	σ_{12}	σ_2^2
M_1, M_4	9/8	0	9/8
M_2	2000	1200	2800
M_3	25	35	14

Table 1: Covariance matrix configurations for models M_1, \dots, M_4 for the extremal spatial dependence analysis.

Tables 2 and 3 summarise the empirical coverage probabilities for nominal 95% confidence intervals for models M_1, \dots, M_4 based on 500 replicate analyses. Columns C_1 provide the interval coverage using the standard Hessian matrix $I(\theta) = H(\theta)$, and columns

C_2 provide the same using the composite likelihood information matrix (4) following Padoan et al. (2010). Columns C_3 correspond to our correction procedure when applied to the intervals in column C_1 using the standard Hessian matrix. For the correction procedure we used $\tilde{\theta} = \hat{\theta}$, the maximum composite likelihood estimate, and $n = 500$ simulated datasets to perform the adjustment.

From Table 2, clearly confidence interval coverage based on the standard Hessian matrix (C_1) is too low. The coverage using the sandwich information matrix (C_2) is very good, with all the reported values close to 0.95. The coverage values obtained using our adjustment procedure, which is based on the intervals in column C_1 , are also very close to 0.95, and mostly closer than with the sandwich information matrix. Similar results are obtained for model M_4 in Table 3. Taken together, these results indicate that our adjustment procedure can successfully modify the upper and lower limits of a confidence interval to achieve comparable results to established methods in complex settings. However, it does not make use of the algebraic representation of $J(\theta)$ in this case, and so is more easily extended to alternative models (e.g. where $J(\theta)$ is not available), albeit at a moderate computational cost.

	M_1			M_2			M_3		
	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
σ_1^2	0.428	0.960	0.960	0.098	0.947	0.950	0.092	0.939	0.944
σ_{12}	0.518	0.940	0.960	0.122	0.955	0.956	0.154	0.921	0.960
σ_2^2	0.468	0.930	0.936	0.092	0.955	0.938	0.102	0.940	0.952

Table 2: Empirical coverage probabilities for 95% confidence intervals of the parameters of models M_1 , M_2 , and M_3 based on 500 replicate analyses. Columns indicate interval confidence estimation methods using: (C_1) the standard Hessian matrix $I(\theta) = H(\theta)$; (C_2) the sandwich information matrix $I(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$; and (C_3) the standard Hessian matrix $I(\theta) = H(\theta)$ followed by our correction procedure.

	σ_1^2	σ_{12}	σ_2^2	α_0	α_1	α_2	β_0	β_1	γ_0
C_1	0.372	0.544	0.388	0.114	0.122	0.098	0.108	0.140	0.104
C_2	0.930	0.925	0.945	0.935	0.945	0.945	0.935	0.940	0.910
C_3	0.924	0.924	0.958	0.950	0.940	0.944	0.944	0.952	0.952

Table 3: Empirical coverage probabilities for 95% confidence intervals of the parameters of model M_4 based on 500 replicate analyses. Columns indicate interval confidence estimation methods using: (C_1) the standard Hessian matrix $I(\theta) = H(\theta)$; (C_2) the sandwich information matrix $I(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$; and (C_3) the standard Hessian matrix $I(\theta) = H(\theta)$ followed by our correction procedure.

3.2 Exchange rate analysis using approximate Bayesian computation

Approximate Bayesian computation (ABC) describes a family of methods of approximating a posterior distribution when the likelihood function is computationally intractable, but where sampling from the likelihood is possible (e.g. Beaumont et al. 2002; Sisson and Fan 2011). These methods can be thought of as constructing a conditional density estimate of the posterior (Blum 2010), where the scale parameter, $h > 0$, of the kernel density

function controls both the level of accuracy of the approximation, and the computation required to construct it. Lower h results in more accurate posterior approximations, but in return requires considerably more computation. As such, moderate values of the scale parameter are often used in practice. Accordingly, this typically results in oversmoothed estimates of the posterior, and in turn, too wide credible intervals.

We consider an analysis of daily exchange rate log returns of the British pound to the Australian dollar between 2005 and 2007. Drovandi and Pettitt (2011) developed an MA(1) type model for these data where the individual log returns were modelled by a g -and- k distribution (Rayner and MacGillivray 2002). The g -and- k distribution is typically defined through it's quantile function

$$Q(z(p); \theta) = a + b \left(1 + c \frac{1 - \exp(-gz(p))}{1 + \exp(-gz(p))} \right) (1 + z(p)^2)^k z(p), \quad (5)$$

where $\theta = (a, b, g, k)$ are parameters controlling location, scale, skewness and kurtosis, and $z(p)$ is the p -quantile of a standard normal distribution. The parameter $c = 0.8$ is typically fixed. We used the sequential Monte Carlo-based ABC algorithm in Drovandi and Pettitt (2011), based on 2,000 particles, to fit the MA(1) model. The data-generation process, used in both ABC and our correction procedure, consists of drawing dependent quantiles $z_i = (\eta_i + \alpha\eta_{i-1})/\sqrt{1 + \alpha^2}$ for $i = 1, \dots, n$, where $\eta_i \sim N(0, 1)$ for $i = 0, \dots, n$, and then substituting $z(p) = z_i$ in (5).

Table 4 shows the estimated 95% central credible intervals, and their widths, for each model parameter based on the ABC kernel scale parameter $h = 0.016$ (following Drovandi and Pettitt 2011) and also the lower value of $h = 0.009$. Also shown are the intervals obtained after performing a local-linear, ridge regression-adjustment (Blum et al. 2012; Beaumont et al. 2002) on the posterior obtained with $h = 0.016$. The regression-adjustment is a standard ABC technique for improving the precision of an ABC posterior approximation, which aims to estimate the posterior at $h = 0$ based on an assumed regression model.

Clearly the parameter credible intervals obtained with $h = 0.009$ are narrower than those obtained with $h = 0.016$, indicating that the larger intervals indeed have greater than 95% coverage. The regression-adjusted intervals generally have widths somewhere between the intervals constructed with $h = 0.016$ and $h = 0.009$. The suggestion from Table 4 is that even narrower (i.e. more accurate) credible intervals may result if it were possible to reduce h further.

Table 5 shows the corrected 95% central credible interval estimates, obtained from the ABC posterior approximations with kernel scale parameter $h = 0.016, 0.02$ and 0.03 . The correction was based on $n = 500$ simulated datasets and using the posterior mean as the estimate $\tilde{\theta}$ of θ . The results of the correction across the three kernel scale parameter values are similar, suggesting potential computational savings in the ABC posterior simulation stage, as one may perform the analysis with larger values of h . All parameters achieve equivalent or improved precision compared to the most precise ABC posterior estimate obtained with $h = 0.009$.

While for a standard Bayesian analysis, the posterior mean is a consistent estimator of θ , this may not be true in the case of the ABC approximate posterior for $h > 0$, as the location and shape of the ABC posterior can change with h . As such, some care may be needed when implementing our correction procedure in this setting. In the current analysis, a preliminary investigation suggested that estimates of the posterior mean stabilised below $h = 0.03$. As such, we are confident that this represents an accurate estimate $\tilde{\theta}$ of θ in this case.

	$h = 0.016$	Width	$h = 0.009$	Width	Reg. Adj. ($h = 0.016$)	Width
a	(-0.0006, 0.0002)	0.0008	(-0.0004, 0.0001)	0.0005	(-0.0006, 0.0002)	0.0008
b	(0.0018, 0.0028)	0.0010	(0.0019, 0.0026)	0.0007	(0.0018, 0.0027)	0.0009
g	(-0.0267, 0.2573)	0.2840	(-0.0044, 0.2138)	0.2182	(-0.0286, 0.2505)	0.2791
k	(0.2024, 0.5061)	0.3037	(0.2607, 0.5322)	0.2715	(0.2148, 0.5092)	0.2944
α	(0.1413, 0.2713)	0.1300	(0.1491, 0.2771)	0.1280	(0.1489, 0.2742)	0.1253

Table 4: 95% central credible intervals and corresponding interval widths from the g -and- k distribution MA(1) model. Results obtained using ABC posterior approximation with kernel scale parameter $h = 0.016$ and $h = 0.009$, and following a ridge regression-adjustment based on an ABC posterior approximation with $h = 0.016$.

	$h = 0.016$	Width	$h = 0.02$	Width	$h = 0.03$	Width
a	(-0.0003, 0.0000)	0.0003	(-0.0003, 0.0000)	0.0003	(-0.0004, -0.0001)	0.0005
b	(0.0020, 0.0024)	0.0004	(0.0021, 0.0025)	0.0004	(0.0021, 0.0024)	0.0003
g	(0.0303, 0.2156)	0.1853	(0.0173, 0.1818)	0.1645	(0.0204, 0.1957)	0.1753
k	(0.2769, 0.4099)	0.1330	(0.2909, 0.4235)	0.1326	(0.2768, 0.4129)	0.1362
α	(0.1430, 0.2659)	0.1229	(0.1335, 0.2708)	0.1373	(0.1513, 0.2714)	0.1201

Table 5: Adjusted 95% central credibility intervals and corresponding interval widths from the g -and- k distribution MA(1) model. Adjusted intervals based on correcting ABC posterior approximations with kernel scale parameter $h = 0.016, 0.02$ and 0.03 .

4 Discussion

In this article we have introduced a method of adjusting confidence interval estimates to have a correct nominal coverage probability. This method was developed in the frequentist framework, but may be equally applied to ensure that Bayesian credible intervals possess the (frequentist) coverage property. Our approach is general and makes minimal assumptions: namely that it is possible to generate data under the same procedure (model) that produced the observed data, and that a consistent estimator is available for the parameter of interest. The correction is asymptotically unbiased, although can work well for moderate sample sizes (m), and there is a central limit theorem for the corrected interval limits in terms of the number (n) of auxiliary samples used to implement the correction.

In the examples that we have considered, we have found that our method can produce confidence intervals which perform comparably to existing gold standard approaches – though with greater scope for extension to more complicated models – and provide a reliable method of adjusting approximately obtained credible intervals in challenging settings.

One potential criticism of our approach is that it requires the construction of a large number (n) of confidence or credible intervals in order to correct one interval. In the case where constructing a single interval is computationally expensive, implementing the correction procedure in full can result in a large amount of computation. This was the case in our exchange rate data analysis using ABC methods, although using an alternative ABC algorithm such as regression-adjustment (based on a single large number of model simulations) would have been more efficient. While regression-adjustment can itself perform poorly if the assumed regression model is incorrect, as our correction procedure makes minimal assumptions, we may still have good confidence, however, in the resulting adjusted intervals it provides.

Acknowledgements

We are grateful to Chris Drovandi and Tony Pettitt for kindly allowing us to use their Matlab codes for the ABC analysis. Financial support for this research was provided by the Australian Research Council Discovery Project scheme (DP1092805) and the University of New South Wales.

Appendix: Proofs

Proof of Theorem 2.1

Writing $L_c(\mathbf{x}) = L(\mathbf{x}) + \delta$, then

$$\begin{aligned} P(\theta < L_c(\mathbf{x})) &= P(\theta < L(\mathbf{x}) + \delta) \\ &= P(\theta - L(\mathbf{x}) < \delta). \end{aligned}$$

Hence, by definition, $P(\theta < L_c(\mathbf{x})) = \alpha/2$ if $\delta = \xi_{\alpha/2}$, where $\xi_{\alpha/2}$ is the $\alpha/2$ -th quantile of the distribution of $\theta - L(\mathbf{x})$.

Proof of Theorem 2.2

Let $G_{\{\theta-L(\mathbf{x})\}}$ be the distribution function of $\theta - L(\mathbf{x})$, which has positive first derivatives so that $G'_{\{\theta-L(\mathbf{x})\}}(\nu) = \frac{\partial}{\partial z} G_{\{\theta-L(\mathbf{x})\}}(\nu) > 0$ for all $\nu \in \mathbb{R}$. Also let $\hat{G}_{\{\tilde{\theta}-L(\mathbf{y})\}}$ be the empirical distribution of $\tilde{\theta} - L(\mathbf{y})$ based on the samples $\tilde{\theta} - L_i(\mathbf{y}_i)$, $i = 1, \dots, n$. From Theorem 2.1 we have that $L_c(\mathbf{x}) = L(\mathbf{x}) + \xi_{\alpha/2}$ for some $\alpha \in (0, 1)$, where $G_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2}) = \alpha/2$. Let $\hat{\xi}_{\alpha/2} = \hat{G}_{\{\tilde{\theta}-L(\mathbf{y})\}}^{-1}(\alpha/2)$ be the empirical estimate of $\xi_{\alpha/2}$.

If we define $\tilde{L}_c(\mathbf{x}) = L(\mathbf{x}) + \hat{\xi}_{\alpha/2}$, then for any $\omega \in \mathbb{R}$, we have

$$\begin{aligned}
\Pr\left(\sqrt{n}(\tilde{L}_c(\mathbf{x}) - L_c(\mathbf{x})) \leq \omega\right) &= \Pr\left(\sqrt{n}(\hat{\xi}_{\alpha/2} - \xi_{\alpha/2}) \leq \omega\right) \\
&= \Pr\left(\hat{\xi}_{\alpha/2} \leq \xi_{\alpha/2} + \omega/\sqrt{n}\right) \\
&= \Pr\left(G_{\{\theta-L(\mathbf{x})\}}(\hat{\xi}_{\alpha/2}) \leq G_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2} + \omega/\sqrt{n})\right) \\
&= \Pr\left(G_{\{\theta-L(\mathbf{x})\}}(\hat{\xi}_{\alpha/2}) \leq \alpha/2 + [\omega G'_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2}) + o(1)]/\sqrt{n}\right)
\end{aligned}$$

where the last equality follows from a first order Taylor expansion of G at $\xi_{\alpha/2}$.

If Y represents the number of times that $G(\hat{\xi}_{\alpha/2})$ is smaller than $\zeta = \alpha/2 + [\omega G'_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2}) + o(1)]/\sqrt{n}$, then since $G(\xi) \sim U(0, 1)$, we have $Y \sim \text{Binomial}(n, \zeta)$. Hence

$$\frac{Y - n\zeta}{\sqrt{n\zeta(1-\zeta)}} \rightarrow N(0, 1) \quad (6)$$

in distribution as $n \rightarrow \infty$ (der Vaart 2000).

Let r_n be the integer rank of the $\alpha/2$ -th quantile from a data set $X = \{X_1, \dots, X_n\}$ of length n , such that $\hat{\xi}_{\alpha/2} = X_{(r_n)}$. If we assume that $\frac{n\alpha/2 - r_n}{\sqrt{n}} \rightarrow 0$ as $n \rightarrow \infty$, then from (6) we have

$$\begin{aligned}
\Pr\left(\sqrt{n}(\hat{\xi}_{\alpha/2} - \xi_{\alpha/2}) \leq \omega\right) &= \Pr\left(G_{\{\theta-L(\mathbf{x})\}}(\hat{\xi}_{\alpha/2}) \leq \zeta\right) = \Pr\left(Y \geq r_n\right) \\
&= \Pr\left(\frac{Y - n\zeta}{\sqrt{n\zeta(1-\zeta)}} \geq \frac{r_n - n\zeta}{\sqrt{n\zeta(1-\zeta)}}\right) \\
&= \Phi\left(\frac{n\alpha/2 + \sqrt{n}\omega G'_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2}) - r_n}{\sqrt{n\zeta(1-\zeta)}}\right) + o_p(1) \\
&= \Phi\left(\frac{n(\alpha/2) - r_n}{\sqrt{n(\alpha/2)(1-\alpha/2)}} + \frac{\omega G'_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2})}{\sqrt{(\alpha/2)(1-\alpha/2)}}\right) + o_p(1) \\
&= \Phi\left(\frac{\omega G'_{\{\theta-L(\mathbf{x})\}}(\xi_{\alpha/2})}{\sqrt{(\alpha/2)(1-\alpha/2)}}\right) + o_p(1). \quad (7)
\end{aligned}$$

It then follows that the consistency of the estimator \tilde{L}_c can be established as

$$\lim_{n \rightarrow \infty} \Pr(\sqrt{n}(\tilde{L}_c(\mathbf{x}) - L_c(\mathbf{x})) \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{E[\sqrt{n}(\tilde{L}_c(\mathbf{x}) - L_c(\mathbf{x}))]}{\epsilon} = 0$$

by the Markov inequality (Ash and Doleans-Dade 2000).

References

- Ash, R. B. and C. Doleans-Dade (2000). *Probability and measure theory*. Academic Press.
- Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics* 162(4), 2025–2035.
- Blum, M. G. B. (2010). Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association* 105, 1178–1187.
- Blum, M. G. B., M. A. Nunes, D. Prangle, and S. A. Sisson (2012). A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science*, in press.
- Cook, S., A. Gelman, and D. Rubin (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* 15, 675–692.
- Davison, A. C. and D. V. Hinkley (1997). *Bootstrap methods and their application*. Cambridge University Press.
- der Vaart, A. W. V. (2000). *Asymptotic statistics*. Cambridge University Press.
- Drovandi, C. C. and A. N. Pettitt (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational Statistics & Data Analysis*.
- Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation (with discussion). *Journal of the Royal Statistical Society: Series B* 74, 419–474.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 221–33.
- Padoan, S. A., M. Ribatet, and S. A. Sisson (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* 105(489), 263 – 277.
- Prangle, D., M. G. B. Blum, G. Popovic, and S. A. Sisson (2012). Diagnostic tools for approximate Bayesian computation using the coverage property. *Tech Report*.
- Rayner, G. D. and H. L. MacGillivray (2002). Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing* 12(1), 57–75.
- Sisson, S. A. and Y. Fan (2011). Likelihood-free Markov chain Monte Carlo. In S. P. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*, pp. 319–341. Chapman and Hall/CRC Press.
- Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5–42.